Niveau(x) d'enseignement: 7 Numéro EDUQ : 8670400
Centre d'intérêt: 570

Gagné, Françoys, 1940-.

The structure of Test PERPE and some psychometric data / by Francoys Gagné with the most helpful collab. of the following members of project PERPE, Denis Allaire... [et al.]. -- [Sainte-Foy] : INRS-Education, 1973. 19, [1] f. ; 28 cm.

"Paper presented as part of a symposium on Teacher Evaluation at the annual meeting of the Canadian Society for the Study of Education, in Kingston, Ontario". Titre de la couv. Code: Document C-07. Bibliogr.: f. [20].

Vedette(s)-secondaire(s): Allaire, Denis • INRS-Education

Descripteur(s): Psychométrie • Elaboration de tests • Evaluation de tests • Instrument d'évaluation • Collégial (Ordre d'enseignement)

Identificateur(s) de nom: PERPE (Questionnaire)

L'objectif principal de cette communication est de faire connaître aux chercheurs en éducation de langue anglaise un instrument pour l'évaluation des enseignants par les étudiants, qui était en 1973 au coeur d'un important projet de recherche au Québec. Il s'agit du Test PERPE (Perceptions étudiantes de la relation professeur-étudiants). Comme cet instrument était presque totalement inconnu en dehors du Québec, le conférencier a cru bon de traiter en détail, dans un premier temps, de la structure même du questionnaire. Dans un second temps, il traite de la fidélité du test. Le texte de la conférence est illustré de onze tableaux.

Type de document: r

EDUQ 8670400

I N R S - E d u c a t i o n

The structure of Test PERPE and some psychometric data

by

François Gagné, Ph. D.
Director of project PERPE

with the most helpful collaboration of
the following members of project PERPE:

Denis Allaire, B.A.

Michel Beaudry, M.A.

Richard Brunet, M.A.

Roger Cormier, Ph.D.

Paper presented as part of a symposium on Teacher Evaluation at the annual meeting of the Canadian Society for the Study of Education, in Kingston, Ontario.

May 1973

Document C-07

A. Introduction

When I started to write the present paper, I considered the possibility of presenting it partly in French, partly in English. Then I remembered what happened in St-John, Newfoundland, in 1970, where we organized a symposium in French on project PERPE. Four persons came, all of them french-speaking. Since I was interested in more people outside of Quebec getting acquainted with our activities, I put aside my nationalist impulses and decided to give this paper in English. My mastery of the english language does not permit me to improvise easily. So you will have to bear a lecture, my accent and my own translation. In order that you may follow more easily, I have made available to you copies of the text.

The first table gives the signification of the PERPE and SPOT abbreviations. The graphic symbol at the top of the table was designed to represent: "the computer serving the relationship between teacher and students". It would be an impossible task to summarize in such a short time the activities we have accomplished in the sphere of project PERPE. The project has been going on for the last six years and we expect to be able to continue our researches and services for many more years. You will find in table 2 a partial list of topics on which I might have dwelt more extensively. It is by no means complete. As project PERPE seems not to be well known outside of Quebec, it was agreed upon with professor Gilles Nadeau that I should focus my paper on the structure of the questionnaire and on some of its psychometric characteristics, that is topics 1, 3 and 4 in table 2.

B. The structure of the questionnaire

The instrument we currently use is composed of 61 items. These were selected from a pool of almost 100 different items that were tried out

Table 1

TEST PERPE

(French title)

Perceptions Étudiantes de la Relation Professeur-Étudiants

(Litteral translation)

Perceptions by Students of the Relationship between Teachers and Students

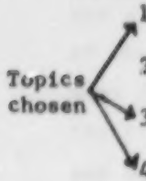
SPOT TEST

Students' Perceptions of Teachers

Table 2

Partial list of topics which could have been included in the paper

RESEARCH



1. The SPOT Test's structure and content.
2. A-priori vs a-posteriori Item groupings.
3. Measures of reliability and Stability.
4. Research on the Validity of the Test.
5. A profile of the typical college teacher in Quebec as seen by students.
6. Relationships between the constructs measured with the SPOT Test.
7. Student characteristics related to their perceptions, expectations and dissatisfactions.
8. Teacher characteristics related to the students' perceptions, expectations and dissatisfactions.
9. Institutional characteristics related to the students' perceptions, expectations and dissatisfactions.
10. Changes in the Teachers' behaviors following the feedback given by students.
11. Construction of a specific instrument for the secondary level.
12. Towards an item bank and a "customized" instrument.

SERVICES

1. The policies and administrative procedures adopted for the test's dissemination.
2. Quality controls and computation flow-chart.
3. The computer profile and Teacher Interpretation Booklet.
4. The decentralized follow-up service.
5. Statistics on the service and the test users.
6. Teacher perceptions of the services offered.

once or more during the experimental phase of the project. More than half of these items were rewritten or rescaled along the road. Table 3 presents a typical item of the SPOT Test. You will note that each item has a title, more or less abstract, followed by an operational description of the trait, attitude or activity we wish to measure. If you wish a more detailed view of the universe of contents covered by the SPOT Test, the titles of the 61 items appear in table 4. At the time the SPOT Test was made available to teachers in french-speaking and english-speaking CEGEPS, we decided to create section scores that would give the teacher a more synthetic view of his profile. The procedure we used starts from the intercorrelation matrix of the items and has been described elsewhere (Gagné, 1970). Recently, we have conducted extensive factor analytic studies of the test which generally confirm this initial grouping of the items, while giving also precise indications as to modifications that could be made to the actual groupings. Table 5 gives the title and a brief description of these six sections.

The items are measured with a five-point Likert-type scale, each point anchored by a quantifying adjective. It seems that this anchoring facilitates choices by the judges.

Finally, using the specific scale of each item, the student answers two questions. First, "Where do you situate this course (this teacher) on the evaluation scale?". Second, "Where should it(he) be situated in order that you be satisfied?". We usually refer to these measures with the terms "reality measure" and "expectations measure".

Compared to all other questionnaires used in north-american colleges and universities, the SPOT Test is the only instrument to our knowledge where dissatisfaction is derived from reality and expectations measures, instead of being either measured directly or being inferred from the 'agree-disagree' type of item. This derivation technique is based on the following postulate:

> "Any discrepancy between the numerical values of an A answer and the corresponding B answer is indicative of dissatisfaction on the part of the student."

Table 3

A typical item in the SPOT Test

05. COURSE STRUCTURE: this concerns the internal organization of the content presented. The presence of such a course structure is evident when the parts of the course seem to be linked one to another, when the professor follows a course outline and when there are frequent subdivisions of subject material.

1. Content STRUCTURED VERY LITTLE
2. Content STRUCTURED LITTLE
3. Content STRUCTURED PARTIALLY
4. Content STRUCTURED
5. Content VERY STRUCTURED

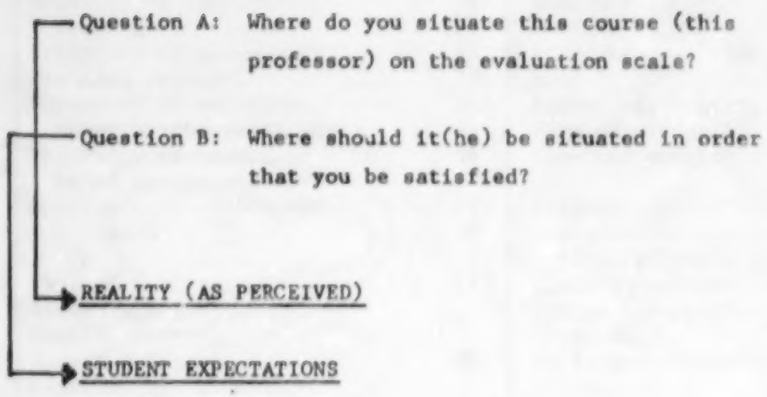

Question A: Where do you situate this course (this professor) on the evaluation scale?
Question B: Where should it(he) be situated in order that you be satisfied?
REALITY (AS PERCEIVED)
STUDENT EXPECTATIONS

Table 4

List of item titles in the current version of the SPOT Test

Item no.	Title	Item no.	Title
01	Quantity of subject matter	31	Captures the attention of the students (content)
02	Fulfillment of expectations	32	Humor
03	Substantial content	33	Dramatic performance
04	Steadiness	34	Intellectually stimulating
05	Course structure	35	Language clarity
06	Information concerning the organization of the content	36	Clarifying value of the examples
07	Syntheses	37	Qualified judgment
08	Planning of the work	38	Value of the professor's questions
09	Frequency of work-verification	39	Ability to clarify
10	Type of questions in work-verification (closed questions)	40	Comprehension of questions
11	Type of questions in work-verification (open-ended questions)	41	Attentive to individual aptitudes
12	Memorization in work-verification	42	Attitude toward individual worth
13	Pedagogical approach	43	Degree of empathy
14	Stability of the pedagogical approach	44	Comprehension in work-verification
15	Fulfillment of expectations	45	Frequency of feedback
16	Desirable digressions	46	Nature of feedback
17	Concrete illustrations	47	Patience
18	Democratic attitude	48	Objectiveness
19	Quantity of compulsory readings	49	Undesirable digressions
20	Frequency of group work	50	Humility
21	Teaching material	51	Justice
22	Frequency of questions asked by the professor	52	Interpersonal atmosphere
23	Frequency of questions asked by the students	53	Mood
24	Frequency of exchanges of ideas	54	Verbal aggressiveness
25	Type of lecture	55	Form of discipline
26	Syntax and vocabulary	56	Expected punctuality
27	General culture	57	Interest in individuals
28	Speech delivery	58	Contacts outside of class time (academic grounds)
29	Captures the attention of the students	59	Contacts outside of class time (non-academic grounds)
30	Captures the attention of the students (tone)	60	Initiative concerning contacts
		61	Nature of exchanges

Table 5

The six sections of the SPOT Test

Title	Description	Number of dimensions
OVERALL STRUCTURE	This section emphasizes the structuring of the learning situation; organisation of the subject matter and techniques of evaluation.	12
VARIETY	This section measures the use, to a greater or lesser extent, of varied teaching techniques.	12
MAGNETISM	This term emphasizes the fact that the evaluation concerns the ability of the teacher to attract and hold the attention of the students through the quality of his teaching.	10
STUDENT-CENTEREDNESS	This group of dimensions evaluates the capacity of the teacher to concentrate on the student, to take into consideration the needs and the limits of a particular group.	12
ATMOSPHERE	The dimensions in this section allow for the evaluation of the socio-emotional aspects of the student-teacher relations within the classroom.	10
AVAILABILITY	This group of dimensions has to do with the number and type of the student-teacher relations outside the classroom.	5

In other terms, we are dissatisfied when things don't work (Reality) as we would like them to (Expectations)! Two corollaries follow from this postulate. First,

> "When the numerical values of an A answer and the corresponding B answer are identical, it is interpreted in termes of student satisfaction."

And secondly,

> "The larger the discrepancy between numerical values of A and B answers, the greater the dissatisfaction."

As we said before, this approach is unique among existing instruments used for the evaluation of teachers by students; however, this idea seems to be gaining more and more attention from researchers. Indeed, two recently published experimental researches, one by Levinthal, Lansky and Andrews (1971), the other by Sanders and Lynch (1973), make use of this derived approach to measure student dissatisfaction. Moreover, as we were reviewing past publications in the field of industrial psychology we came upon a work satisfaction questionnaire, built on premises quite identical to ours; it is the PNSQ, that is Porter's Need Satisfaction Questionnaire (Porter, 1961; Porter, 1962). Many researches have been conducted using this instrument (Cummings and ElSalmi, 1968). But, we know of only one study (Imparato, 1972) that attempts to validate this derived approach.

From our two direct measures, we derive three dissatisfaction scores. As shown in table 6, we first compute the percentage of dissatisfied students, those whose A answer differs from the corresponding B answer. Secondly, we calculate a GROSS dissatisfaction score: it is the sum of the differences <u>in absolute values</u>, divided by the total number of students answering (satisfied <u>and</u> dissatisfied). Thirdly, we compute a NET dissatisfaction score: we use the algebraic sum of the differences between reality and expectations measures. We usually represent these three scores by the following symbols: %, MAB and MAL.

Table 6

A. Three measures of dissatisfaction

Symbol	Definition	Formula
%	Percentage of dissatisfied students	$\frac{N(A \neq B) \times 100}{N}$
MAB	GROSS measure of dissatisfaction (Mean of ABsolute discrepancies)	$\frac{\Sigma(\lvert A - B \rvert)}{N}$
MAL	NET measure of dissatisfaction (Mean of ALgebric discrepancies)	$\frac{\Sigma(A - B)}{N}$
PCD	Percentage of intra-group contradiction	$\frac{(MAB - \lvert MAL \rvert) \times 100}{MAB}$

B. Example, using item 47: Patience

Group X	Group Y
MA = 3.20 (OFTEN)	MA = 4.00 (VERY OFTEN)
MB = 4.10 (VERY OFTEN)	MB = 3.80 (VERY OFTEN)
% = 56	% = 36
MAB = 0.90	MAB = 0.60
MAL = -0.90	MAL = +0.20
PCD = 0%	PCD = 67%

The first score, the percentage of dissatisfied students, lends itself to a straightforward interpretation. The two other scores, MAB and MAL, are not interpreted as easily; they usually take their full meaning when related one to another. Let's clarify their respective meaning with the help of two fictitious examples using item 47: the teacher's patience. This item's description insists particularly on the teacher's willingness to repeat explanations for those students who have not understood the first time. The five-point scale starts with "RARELY patient" and ends with "ALWAYS patient". As shown in table 6, group X gives a reality score of 3.20, which corresponds to "OFTEN patient", while the mean score for expectations is 4.10, that is "VERY OFTEN patient". 56% of the students in group X were judged dissatisfied; and because the discrepancy varies in degree, the gross dissatisfaction score came up to 0.90. Moreover, we note an MAL that is negative in sign and equal in value to the MAB. The identity between these two scores indicates that all dissatisfied students gave the same direction to their dissatisfaction, that is they all expected their teacher to be more patient than he was. On the other hand, the mean of A answers in group Y is much higher, 4.00, while the MB appears slightly lower, that is 3.80. Here, 36% of the students have given a B answer that differs numerically from their A answer. The MAB or gross dissatisfaction score is equal to .60, but the MAL differs considerably, being at +0.20. The positive sign of the MAL shows that the majority of dissatisfied students want the teacher to be <u>less</u> patient than he is. However, the difference in value between the two dissatisfaction scores shows ample disagreement among the dissatisfied students as to the direction of their dissatisfaction: some would like even more patience, while others (the majority) find it already too high.

Observations like these have lead us to a fourth construct: the percentage of intra-group contradiction, which we symbolize by the letters PCD. The formula by which we compute PCD is shown in table 6. When all differences are of a same sign, PCD will equal 0, while an equal sum of positive and negative values will give a PCD equal to 100%.

You will note that our derived approach alone ables us to make allowance for the direction of the dissatisfaction expressed by students. Moreover, only this approach makes possible the calculation of intra-group contradiction. This phenomenon appears to us to be quite interesting. Indeed, the teacher's patience is spontaneously considered a highly desirable quality. In a conventional questionnaire where the student will be asked directly his degree of satisfaction as to the teacher's patience, the presence of any dissatisfaction on the part of the student may be automatically interpreted in terms of lack instead of excess. Yet, we have discovered that excesses are by no means rare.

Table 7 shows a correlation diagram in which the computer plotted 800 classes in terms of the MAs and MBs obtained by teachers. The diagonal through the diagram connects all points where MA = MB, that is where the net dissatisfaction score is zero. Classes in the upper triangle have signaled a deficiency as to the teacher's patience: MB is higher than MA. In the lower triangle we find the excess cases. Their number is by no means negligeable. So, even though patience is considered a highly desirable quality on the part of the teacher, as witnessed by the generally high level of the MBs in the diagram, in many instances this quality manifests itself excessively. This phenomenon brings to my mind a well-known english saying: "Too much of a good thing is good for nothing".

C. A few psychometric data

The field of psychometric research is usually subdivided in two partially related sectors: reliability and validity. Because of the time limit imposed, I realized that I could not cover the two sectors adequately; my paper was already too long. As we have almost completed our main reliability studies, I decided to focus on this aspect.

In the last few years, the concept of reliability has been the object of important clarifications, thanks in good part to the contribu-

DIAGRAMME DE CORRELATION.

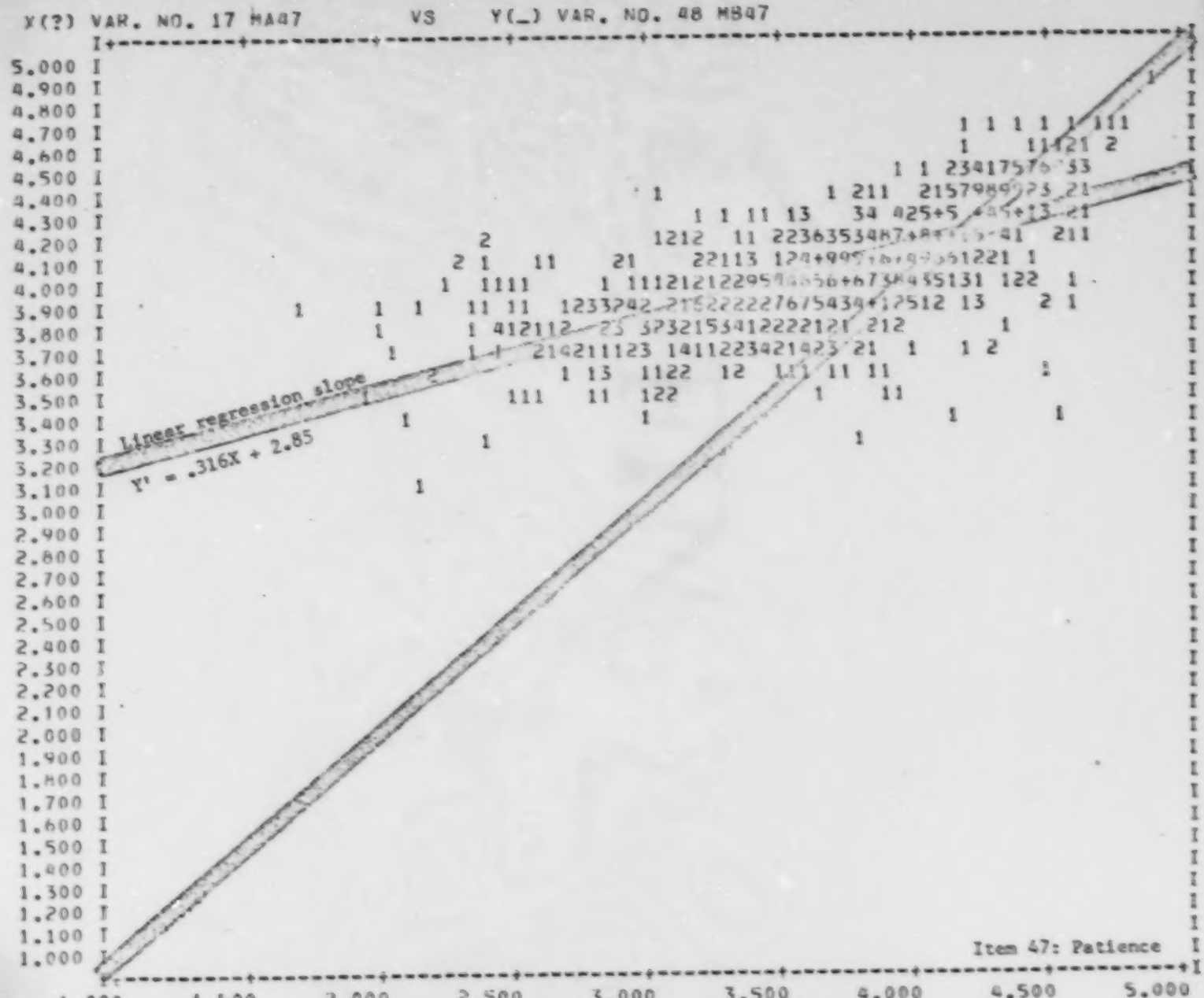


Table 7

Correlation diagram of 800 classes in terms of the MAs and MBs obtained by teachers on item 47: Patience

tions of analysis of variance techniques. The chapter by Stanley (1971) in Thorndike's Educational Measurement, the revised Standards for Development and Use of Educational and Psychological Tests, sponsored conjointly by APA, AERA and NCME, and the book by Cronbach and others (1972), all of these demonstrate the conceptual refinements brought to this construct. Essentially, we no more speak of THE reliability of a measuring instrument, but of different reliability coefficients depending upon the sources of variation we include in the "true" variance or in the error variance.

This problem, highly technical at first sight, has nevertheless important practical implications. On the one hand, the typical user does not always resist the temptation to generalize in time and space the results obtained at one particular moment, with a specific group and a specific course. He will say: "I received a low score; then I must be a poor teacher". On the other hand, the reliability coefficients commonly published, be they split-halves, analysis of variance estimates, or others, usually do not scientifically permit any generalization beyond a specific teaching context. We have tried to relate our reliability studies to this methodological framework. Practically, our technical problems are quite complex, since we have to analyze individually the sixty-one items of the test crossed with the five structural constructs, plus the six section scores.

A first reliability study was concerned with the individual items. We used a formula developed by Ebel (1951): it corresponds to an analysis of variance where each examinee is evaluated by a different group of judges. This coefficient defines as "true score" the score a teacher would obtain if an infinite number of comparable student groups had gone through an identical learning situation and had evaluated him at the same moment. Here, the generalization bears upon the judges but ignores certain circumstantial influences, like the specific subject matter teached or the time of the evaluation, which remain undissociated from the variance due to the teachers. On the other hand, the judges' inaccuracy, their varying degree of leniency and their different points of view as to what makes a teacher satisfying, all of these contribute to the error variance.

Table 8

Reliability of the judges for the 61 items and five main constructs of the SPOT Test

Item	MA	MB	MAL	MAB	%
01	.94	.84	.88	.80	.76
02	.82	.62	.67	.76	.70
03	.85	.59	.78	.84	.76
04	.85	.53	.81	.82	.73
05	.90	.75	.84	.85	.78
06	.86	.59	.81	.82	.76
07	.80	.56	.75	.77	.68
08	.82	.63	.75	.75	.69
09	.94	.88	.85	.76	.69
10	.91	.84	.71	.68	.62
11	.85	.76	.64	.62	.58
12	.90	.81	.77	.73	.68
13	.90	.80	.77	.74	.68
14	.89	.72	.80	.80	.71
15	.74	.54	.59	.74	.69
16	.91	.69	.84	.84	.78
17	.88	.73	.78	.77	.71
18	.91	.78	.85	.82	.76
19	.95	.88	.86	.66	.60
20	.97	.90	.91	.84	.79
21	.98	.94	.93	.93	.90
22	.90	.70	.82	.79	.71
23	.88	.62	.81	.79	.71
24	.92	.84	.82	.80	.74
25	.90	.76	.83	.85	.79
26	.92	.79	.83	.77	.72
27	.89	.66	.80	.80	.74
28	.91	.45	.90	.88	.84
29	.95	.59	.93	.93	.89
30	.91	.58	.88	.88	.82
31	.94	.49	.92	.92	.87
32	.95	.73	.90	.89	.86
33	.93	.73	.87	.87	.81
34	.88	.61	.83	.84	.75
35	.87	.49	.82	.83	.74
36	.86	.54	.84	.84	.77
37	.81	.60	.70	.77	.67
38	.90	.64	.84	.84	.75
39	.87	.51	.86	.86	.77
40	.85	.50	.82	.82	.70
41	.84	.54	.79	.79	.71
42	.86	.59	.79	.80	.73
43	.87	.48	.83	.84	.74
44	.81	.55	.71	.74	.64

Table 8 (continued)

Reliability of the judges for the 61 items and five main constructs of the SPOT Test

Item	MA	MB	MAL	MAB	%
45	.88	.70	.76	.74	.66
46	.82	.55	.73	.76	.66
47	.89	.64	.84	.82	.76
48	.78	.43	.70	.75	.64
49	.84	.57	.77	.80	.73
50	.87	.49	.84	.84	.79
51	.81	.51	.76	.77	.67
52	.78	.38	.80	.87	.81
53	.91	.72	.87	.87	.82
54	.88	.72	.75	.78	.71
55	.87	.73	.81	.77	.69
56	.93	.82	.81	.61	.53
57	.89	.66	.83	.82	.77
58	.87	.70	.77	.75	.71
59	.87	.66	.65	.65	.62
60	.87	.62	.78	.76	.71
61	.87	.62	.78	.76	.71
Mean	.90	.68	.81	.80	.74

Table 8 lists the reliability coefficients obtained for each item crossed with the five structural constructs. They were computed on a sample of 1200 classes, averaging 27 students. To simplify the analysis we have classified them into six categories, as shown in table 9. Here are a few general observations concerning these results.

a) Item after item, the highest coefficients are obtained with the reality measures.

b) We have not been much surprised by the relatively low coefficients of the MBs or expectations measures. They are not alarming since we do not use them to discriminate between teachers. Moreover, we do not expect to get coefficients that would be much higher; the standard deviations of the MBs are small and cannot be much enlarged, since most of the scales

Table 9

Distribution of the 61 items into six categories in terms of their reliabilities (as given in table 8)

Categories	MA	MB	MAL	MAB	%
.90+	24	2	6	3	1
.85 - .89	25	2	8	9	3
.80 - .84	9	6	20	21	5
.75 - .79	2	5	19	17	15
.70 - .74	1	10	4	6	18
.69-	-	36	4	5	19
Totals	61	61	61	61	61

are value loaded, so that expectations measures tend to concentrate near points 1 or 5.

c) The inaccuracy of the expectations measures affects the reliability coefficients of the MAB and MAL measures. Yet, their level compares favorably to coefficients given in the litterature, as far as we can judge from the incomplete information given in most articles reviewed (Costin, 1968; Gage and Suci, 1952; Remmers, 1934; Spencer and Aleamoni, 1970).

Besides this first study, we also asked ourselves whether certain characteristics of the students could be associated with some systematic biases in the measures. We introduced the following variables in a multiple regression model:

> the student's sex, age, level, stream (general or technical), program of studies (physical sciences, biological sciences, human sciences, arts and letters), standardized academic standing in the course evaluated, socio-economic level, the status of the course taken (optional or compulsory), and finally the fact he was repeating or not the course.

Table 10

Average over the 61 items of two different reliability coefficients for each of four constructs

Type of coefficient	MA	MB	MAL	MAB
"Between-judges" reliabilities (ANOVA):	.90	.68	.81	.80
Between pairs correlations:	.68	.47	.59	.57

The results, which we expect to publish soon, show that none of these variables account for more than 5% of the variance of the student answers. These findings agree with most published reports on student variables (cf. Costin, Greenough and Menges, 1971).

In the same context, one of us is working on a project to identify reliability coefficients adapted to larger universes of generalization. He wanted to know particularly to what extent a profile of results remains stable no matter which specific course he gives or to which specific group he addresses himself. This type of generalization is made possible by the fact that many users administer the test to two groups inside the same semester. By computing item by item the correlation coefficients between pairs of classes and then comparing those to the reliability coefficients obtained for a more restrictive universe of generalization (for example the coefficients stated earlier in this paper), we obtain a "situational stability index" for each item. Some preliminary results of this study appear in table 10. The correlation coefficients were computed on a sample of 149 pairs of classes, selected in the same semester. These data clearly show a distinct drop in the reliability coefficients, when we transfer from the "true" variance to the error variance certain sources of variation pertaining to the teaching context. It will be quite interesting to identify which items appear more (or less) affected by these situational factors.

Table 11

Two reliability coefficients of the section scores

Sections	Number of items	MAL		MAB	
		alpha	Split-half	alpha	Split-half
1	12	93	80	93	80
2	12	92	84	92	82
3	10	90	85	89	86
4	12	93	73	93	76
5	10	90	84	91	66
6	5	81	77	81	77
Global score	61	99	-	99	-

To conclude this overview of our reliability studies, I will summarize briefly our results on the reliability of the section scores. At this level, only the dissatisfaction scores are considered when we apply a simple averaging procedure to compute the section scores. We used two different techniques to estimate their reliability. On the one hand, Cronbach's alpha coefficient (cf. Stanley, 1971, p. 411) gives an estimate of the representativity of the items as measures of an underlying construct; in other terms, it indicates to what degree we can generalize to the universe of items similar to those that make up the sections. The comparison of section scores made from split-halved groups of students uses the same definition of error variance as in the first study summarized above; this technique will show how well we can generalize to the universe of groups of judges comparable to those that were actually used.

Table 11 gives the two reliability coefficients obtained for each section and for the global dissatisfaction score. We considered these coefficients quite satisfactory. So, even though the standard deviations of the section scores are much smaller than those of the items, they offer as good a discrimination index as the item scores.

D. Summing up

The main objective of this paper was to make known to the english community of educational researchers an instrument for the evaluation of teachers by students, which is at the heart of an important project of research and services in Quebec. As this project seemed almost unknown outside of our province, we deemed necessary to detail the instrument's format. Then, in the time left, we decided to focus on our reliability studies rather than give a brief overview of both fields, that is reliability and validity. As we say in French: "Qui trop embrasse, mal étreint! (Grasp all, lose all)". There sure will be an other occasion like this one where we will be able to discuss more extensively some problems related to validity, a domain in which we already have gathered some interesting data.

References

COSTIN, F. (1968). A graduate course in the teaching of psychology: description and evaluation. J. Teacher Education, 19, 425-432.

COSTIN, P., GREENOUGH, W.T. & MENGES, R.J. (1971). Student ratings of college teaching: reliability, validity and usefulness. Review of Educational Research, 41, 5, 511-535.

CRONBACH, L.J., GLESER, G.C., NANDA, H. and RAJARATNAM, N. (1972). The dependability of behavioral measurements. New York: Wiley.

CUMMINGS, L.L. & ELSALMI, S.M. (1968). Empirical research on the bases and correlates of managerial motivation: a review of the litterature. Psychol. Bulletin, 70, 127-144.

EBEL, R.L. (1951). Estimation of the reliability of ratings. Psychometrika, 16, 4, 407-424.

GAGE, N.L. & SUCI, G. (1952). Social perceptions and teacher-pupil relationship. J. Educ. Psychol., 42, 144-152.

GAGNE, F. (1970). Les six nouvelles sections du test PERPE. PERPE, document 0670-01.

IMPARATO, N. (1972). Relationship between Porter's Need Satisfaction Questionnaire and the Job Descriptive Index. J. of Appl. Psychol., 56, 5, 397-405.

LEVINTHAL, C.F., LANSKY, L.M. & ANDREWS, D.E. (1971). Student evaluations of teacher behaviors as estimations of real-ideal discrepancies: a critique of teacher rating methods. J. Ed. Psychol., 62, 2, 104-109.

PORTER, L.W. (1961). A study of perceived need satisfactions in bottom and middle management jobs. J. Appl. Psychol., 45, 1-10.

PORTER, L.W. (1962). Job attitudes in management: I. Perceived deficiences in need fulfillment as a function of job level. J. Appl. Psychol., 46, 6, 375-384.

REMMERS, H.H. (1934). Reliability and halo effects of High School and College students' judgments of their teachers. J. Appl. Psychol., 18, 619-630.

SANDERS, J.R. & LYNCH, M. (1973). Student evaluation of instruction: the analysis of discrepancies between perceived and ideal conditions. Research paper no. 8. Bloomington, Indiana. Educational Research and Evaluation Laboratory, Indiana University, 20 p.

SPENCER, R.E. & ALEAMONI, L.M. (1970). A student course evaluation questionnaire. J. Educ. Measurement, 7, 3, 209-210.

STANLEY, J.C. (1971). Reliability. In Thorndike, R.L., Educational measurement (2nd ed.). Washington: American Council on Education, 356-442.